

REMARKS

I. Status of the Claims

Claims 1, 2, 5-10, 12, 13, 15, 17-29 are currently pending under examination. Upon entry of the present amendment, claim 1 is amended to insert "that is" in line 2 before "at least 60% identical to SEQ ID NO:4," "a" in line 5 before "promoter, and "endogenous *FIE*" in line 5 between "inhibit" and "gene expression." Support for inhibition of the endogenous *FIE* gene expression is found in the specification, *e.g.*, on page 15, lines 5-27. Claim 2 is amended to delete the word "about." No new matter is introduced by the present amendment.

II. Claim Objections

Claim 1 was objected to in the final Office Action of July 31, 2003, for omitting an amendment purportedly made and for omitting the article "a" before the word "promoter." These informalities have been cured by the present amendment.

III. Double Patenting

The Examiner rejected pending claims under judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-6, 9-17, and 20-22 of U.S. Patent No. 6,229,064.

Applicants submit with this Amendment a terminal disclaimer. Applicants note that the filing of a terminal disclaimer to obviate a rejection based on non-statutory double patenting is not an admission of the propriety of the rejection. *See*, MPEP §804.02. Accordingly, Applicants respectfully request withdrawal of the rejection.

IV. Claim Rejections

A. 35 U.S.C. §112, Second Paragraph

Claims 2 and 15-29 were rejected under 35 U.S.C. §112, second paragraph, for alleged indefiniteness. Specifically, the Examiner stated that the word "about" in claim 2 should be deleted but was not properly deleted in Applicants' last response. The Examiner further pointed to the phrase "modulating endosperm development" in claim 15, alleging that the phrase

creates inconsistency with claim 1, from which claim 15 depends, because claim 1 recites "enhances endosperm development."

As amended, the word "about" is deleted from claim 2. The phrase "modulating endosperm development" in claim 15 is amended to "enhancing endosperm development." The indefiniteness rejection is thus obviated.

B. 35 U.S.C. §112, First Paragraph: Written Description

Claims 1, 2, 5-7, 9, 10, 12, 13, 15, 18, and 20-28 were rejected under 35 U.S.C. §112, first paragraph, for alleged failure to meet the written description requirement. As amended, independent claim 1 is drawn to a nucleic acid comprising a *FIE* polynucleotide encoding a polypeptide that is at least 60% identical to SEQ ID NO:4 and comprises a WD40 repeat. The claimed nucleic acid is further defined by its capability of enhancing endosperm development in the absence of fertilization when the polynucleotide is operably linked to a promoter to inhibit endogenous FIE gene expression and introduced into a plant.

The Examiner asserted that the WD40 repeat is present in proteins that have functions different from the FIE polypeptides of the present invention and that the presence of a WD40 repeat therefore does not correlate a structure with the functionality of a FIE polypeptide. The Examiner further asserted that the claimed subject matter is not fully described because no description is provided for a polypeptide that has 60% identity to SEQ ID NO:4 and retains the functionality of SEQ ID NO:4. Applicants respectfully traverse the rejection.

As discussed in Applicants' previous communication to the PTO filed May 15, 2003, the pending claims of the present application fully comply with the written description requirement as set forth in *University of California v. Eli Lilly & Co.*, 43 USPQ2d 1398 (Fed. Cir. 1997) by providing both the shared structural feature (encoding a polypeptide that has at least 60% sequence identity to SEQ ID NO:4 and has a WD40 repeat) and functional feature (capable of enhancing endosperm development in the absence of fertilization in a plant when introduced under a promoter) of the claimed nucleic acids. While the presence of a WD40 repeat alone may not confer functionality, its combination with the specified percent identity does.

To address the Examiner's specific concerns regarding the existence of polypeptides that are at least 60% identical to SEQ ID NO:4 yet retain the desired functionality, Applicants hereby submit a reference by Springer *et al.*, *Plant Physiology* 128:1332-1345 (2002) (Exhibit A), which describes the structure of two maize genes, *ZmFie1* and *ZmFie2*. As their names reflect, Exhibit A identifies *ZmFie1* and *ZmFie2* as homologs of Arabidopsis *FIE*, the subject of the present application. *See, e.g.*, Springer *et al.*, page 1336, right column.

Applicants have also provided sequence alignments between SEQ ID NO:4 and the *ZmFie1* and *ZmFie2* polypeptides (Exhibit B). The alignments, prepared with BLAST, indicate that there is approximately 68% and 58% identity, respectively, with *FIE*. Therefore, Springer *et al.* provide evidence that sequences with identity as low as 58% identity (i.e., very close to, but below, the claimed 60%) are homologs of *FIE*.

Springer *et al.* thus provides evidence that the guidance in the specification is sufficient for those of skill in the art to identify gene products within the scope of the present claims. The present application teaches that *FIE* polypeptides are related to well-known *Drosophila* polycomb proteins comprising WD40 repeats, including "*extra sex combs*" (*esc*). *See*, page 30, lines 16-18 and page 4, first full paragraph of the specification. Springer *et al.* relies on these same relationships with *Drosophila* proteins to identify *ZmFie1* and *ZmFie2*. Thus, Springer *et al.* is in fact evidence that the structures described in the present application adequately described other sequences within the scope of the claim such that those of skill in the art could identify them. Thus, it is clear that the inventors were in possession of the presently claimed genus of polynucleotides.

In light of the fact that (1) the gene structure of *FIE* homologs such as fruit fly *Esc* genes was known and described, and (2) homologs of *FIE* with the same *Esc*-like gene structure have amino acid sequence identities as low as 58% to the *FIE* polypeptide, Applicants submit that the present disclosure reasonably conveys to one of ordinary skill in the art that the present inventors had the claimed subject matter in their possession as of the filing

date. Accordingly, Applicants thus respectfully request the withdrawal of the written description rejection.

C. 35 U.S.C. §112, First Paragraph: Enablement

Claims 1, 2, 5-7, 10, 12, 13, 15, 18, and 20-28 were also rejected under 35 U.S.C. §112, first paragraph, for alleged failure to meet the enablement requirement. The Examiner specifically asserted that insufficient guidance is provided for modifying the amino acid sequence of SEQ ID NO:4 without altering its functionality, particularly because the presence of a WD40 repeat alone does not ensure the retention of the functionality and because sequences with as much as 40% sequence divergence were allegedly not described. Applicants respectfully traverse the rejection.

To establish a *prima facie* case of non-enablement, the Examiner must show that undue experimentation would be required to make and use the claimed invention. Even if the practice of the claimed invention requires a considerable amount of experimentation, it is not necessarily “undue” experimentation:

The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed. *In re Wands*, 8 USPQ2d 1400 (Fed. Cir. 1988) (citing *In re Angstadt*, 190 USPQ 214 (CCPA 1976).
MPEP § 2164.06.

The thrust of the rejection’s argument is that it would require undue experimentation to identify FIE sequences with at least 40% sequence divergence from SEQ ID NO:4. As described above, the specification provided sufficient disclosure regarding specific motifs (e.g., WD40 motifs) and relationships with known proteins (e.g., Esc) for those of skill in the art to make and use sequences within the full scope of the claims without undue experimentation. Indeed, using information provided in the present specification regarding the relationship with *Drosophila polycomb* proteins, Springer *et al.* identified FIE homologs with even less than the 40% divergence allowed for in the present claims. This demonstrates that

Appl. No. 09/812,283
Amdt. dated January 30, 2004
Amendment under 37 CFR 1.116 Expedited Procedure
Examining Group

PATENT

those of skill in the art were fully enabled to practice the claimed invention within the full scope of the claims.

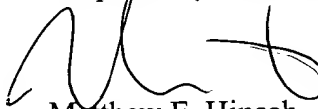
Accordingly, Applicants thus request that the Examiner withdraw the enablement rejection.

CONCLUSION

In view of the foregoing, Applicants believe all claims now pending in this Application are in condition for allowance and an action to that end is respectfully requested.

If the Examiner believes a telephone conference would expedite prosecution of this application, please telephone the undersigned at (415) 576-0200.

Respectfully submitted,



Matthew E. Hinsch
Reg. No. 47,651

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, Eighth Floor
San Francisco, California 94111-3834
Tel: 415-576-0200
Fax: 415-576-0300
Attachment (Exhibit A: Springer *et al.*; Exhibit B: results of sequence alignment)
MEH:cg
60117180 v1

Sequence Relationships, Conserved Domains, and Expression Patterns for Maize Homologs of the Polycomb Group Genes *E(z)*, *esc*, and *E(Pc)*¹

Nathan M. Springer, Olga N. Danilevskaya, Pedro Hermon, Tim G. Helentjaris, Ronald L. Phillips, Heidi F. Kaeppler, and Shawn M. Kaeppler*

Department of Agronomy, University of Wisconsin, Madison, Wisconsin 53706 (N.M.S., H.F.K., S.M.K.); Pioneer Hi-Bred International, Johnston, Iowa 50131 (O.N.D., P.H., T.G.H.); and Department of Agronomy, University of Minnesota, St. Paul, Minnesota 55108 (R.L.P.)

Polycomb group (PcG) proteins play an important role in developmental and epigenetic regulation of gene expression in fruit fly (*Drosophila melanogaster*) and mammals. Recent evidence has shown that Arabidopsis homologs of PcG proteins are also important for the regulation of plant development. The objective of this study was to characterize the PcG homologs in maize (*Zea mays*). The 11 cloned PcG proteins from fruit fly and the *Enhancer of zeste* [*E(z)*], *extra sex combs* (*esc*), and *Enhancer of Polycomb* [*E(Pc)*] homologs from Arabidopsis were used as queries to perform TBLASTN searches against the public maize expressed sequence tag database and the Pioneer Hi-Bred database. Maize homologs were found for *E(z)*, *esc*, and *E(Pc)*, but not for *Polycomb*, *pleiohomeotic*, *Posterior sex combs*, *Polycomblike*, *Additional sex combs*, *Sex combs on midleg*, *polyhomeotic*, or *multi sex combs*. Transcripts of the three maize *Enhancer of zeste*-like genes, *Mez1*, *Mez2*, and *Mez3*, were detected in all tissues tested, and the *Mez2* transcript is alternatively spliced in a tissue-dependent pattern. *Zea mays* fertilization independent endosperm1 (*ZmFie1*) expression was limited to developing embryos and endosperms, whereas *ZmFie2* expression was found throughout plant development. The conservation of *E(z)* and *esc* homologs across kingdoms indicates that these genes likely play a conserved role in repressing gene expression.

Gene expression patterns in eukaryotes are regulated in response to developmental and environmental cues. Changes in the patterns of gene expression are often the result of specific transcriptional regulators. In many cases, patterns of gene expression must be stably maintained through mitotic cell divisions even though the transcriptional regulator that effected the change in expression is only present transiently. The Polycomb group (PcG) genes of fruit fly (*Drosophila melanogaster*) stabilize repressed chromatin states during development. Recently, homologs of PcG genes have also been shown to affect developmental gene regulation in other species.

Simon (1995) defined a set of 13 PcG genes in fruit fly based on a common phenotype of homeotic transformation. The homeotic transformations caused by mutations in PcG genes are the result of a failure to maintain transcriptional repression of homeotic genes through development. Biochemical and genetic evidence indicates that the 13 PcG proteins operate in at least two distinct complexes (Franke et al., 1992; Strutt and Paro, 1997; Jones et al., 1998; Sewalt et al.,

1998; van Lohuizen et al., 1998; Ng et al., 2000; Tie et al., 2001). One complex includes the PcG proteins *E(Z)* and *ESC*, as well as the histone deacetylase *RPD3* and the histone-binding *p55* proteins (Tie et al., 2001). The second complex includes *Polycomb* (*PC*), *Posterior sex combs* (*PSC*), *Polyhomeotic* (*PH*), *Sex combs on Midleg* (*SCM*), and additional uncharacterized proteins (Franke et al., 1992; Shao et al., 1999; Poux et al., 2001). PcG proteins have also been shown to repress expression of introduced (Pal-Bhadra et al., 1997, 1999) and endogenous (Laible et al., 1997) genes in fruit fly. All examples of polycomb-based repression likely operate through formation of a repressive chromatin structure.

Mammalian homologs of all of the cloned PcG proteins, except *multisex combs*, have been documented (Simon, 1995; Schumacher and Magnuson, 1997; Brock and van Lohuizen, 2001). As in fruit fly, mutations in mammalian PcG genes result in anterior derepression of *Hox* gene expression and alterations in cellular proliferation patterns (van der Lugt et al., 1994; Alkema et al., 1995; Muller et al., 1995; Akasaka et al., 1996; Core et al., 1997; Gould, 1997). In *Caenorhabditis elegans*, homologs of three of the 11 cloned PcG proteins, *Enhancer of zeste* [*E(z)*], *extra sex combs* (*esc*), and *Enhancer of Polycomb* [*E(Pc)*], have been reported (Holdeman et al., 1998; Korf et al., 1998; Stankunas et al., 1998). The *E(z)* and *esc* homologs (*maternal effect sterile-2* [*mes-2*] and *maternal effect sterile-6* [*mes-6*]) from *C. elegans* were identified as

¹ This research was supported by the U.S. Department of Agriculture (National Needs Fellowship no. 98-38420-5832) and by the National Science Foundation (grant no. 9975930).

* Corresponding author; e-mail smkaeppl@facstaff.wisc.edu; fax 608-262-5217.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.010742.

maternal genes required for germline immortality (Holdeman et al., 1998; Korf et al., 1998). The *mes-2* and *mes-6* genes are also involved in the silencing of transgenes in germline cells (Kelly and Fire, 1998).

Homologs of *E(z)* and *esc* have also been documented in Arabidopsis (Goodrich et al., 1997; Grossniklaus et al., 1998; Ohad et al., 1999). Three *E(z)*-like genes, *curly leaf* (*clf*; Goodrich et al., 1997), *Medea* (*Mea*; Grossniklaus et al., 1998), and *E(z)-likeA1* (*Eza1*; Preuss, 1999) and one *esc*-like gene, *fertilization-independent endosperm* (*fie*; Ohad et al., 1999), have been previously described.

Mea (also identified as *emb173*, *fis1*, and *f644*) was identified in a screen for Arabidopsis gametophyte lethal mutations and autonomous endosperm development mutations (Castle et al., 1993; Chaudhury et al., 1997; Grossniklaus et al., 1998; Kiyosue et al., 1999; Luo et al., 1999). A plant heterozygous for a *mea* mutation will produce 50% aborted seeds that collapse and do not germinate. It has subsequently been found that *Mea* is imprinted such that the maternal copy of *Mea* is expressed endosperm tissues, whereas the paternal copy is not (Kinoshita et al., 1999; Vielle-Calzada et al., 1999). *Mea* mutants fail to repress endosperm development in the absence of fertilization (Kiyosue et al., 1999). These results indicate that maternal expression of *Mea* is required to repress endosperm development.

Plants homozygous for *clf* mutations display curled leaves, altered maturation times, and partial homeotic transformations of floral tissues (Goodrich et al., 1997). Ectopic expression of the homeotic genes *Agamous* (*AG*) and *Apetala3* (*AP3*) is also observed in *clf* homozygotes. In plants with wild-type *Clf* alleles, *AG* and *AP3* are expressed in floral tissues where *Clf* mRNA is present. The presence of *Clf* RNA in cells expressing *AG* and *AP3* indicates that CLF protein is not sufficient to organize suppressive chromatin, but that other targeting factors are also required (Goodrich et al., 1997). Overlapping expression of PcG and homeotic genes also occurs in fruit fly. The coexpression of functional PcG proteins and the genes that they can repress indicates that specific targeting factors are involved in PcG-dependent repression. A third *E(z)*-like gene, *Eza1*, is present in the Arabidopsis genome (Preuss, 1999). No phenotype for *Eza1* has been reported.

Mutations in the Arabidopsis *esc*-like gene, *fie*, produce phenotypes very similar to *Mea* (Ohad et al., 1996). A female gametophyte with a *fie* mutant allele will undergo replication of the central cell nucleus and partial endosperm development without a fertilization event (Ohad et al., 1999). This indicates that FIE is involved in repressing endosperm development until fertilization occurs. The similar phenotypes of *fie* and *mea* mutants suggests that these two genes may participate together in a silencing complex. This is consistent with the proven direct interaction of E(Z) and ESC in fruit fly (Jones et al., 1998),

an interaction also shown to occur between FIE and MEDEA (Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000).

The objective of this research was to characterize the PcG genes (defined by Simon, 1995) in maize. Fruit fly and Arabidopsis sequences were used to identify maize sequences homologous to PcG genes. We report the full-length cDNA sequence of three *E(z)* homologs, two *esc* homologs, and one *E(Pc)* homolog, and we describe the likely evolution of these sequences. We also describe expression patterns of these genes in maize.

RESULTS

The 11 cloned (two other PcG genes have not been cloned) fruit fly proteins defined as the PcG by Simon (1995) were used as queries to identify homologs in Arabidopsis and maize (Table I). The top TBLASTN score from searches of the complete Arabidopsis genome sequence is shown for each PcG protein. The existence of *E(z)* and *esc* homologs in Arabidopsis has been previously documented (Goodrich et al., 1997; Grossniklaus et al., 1998; Ohad et al., 1999). The BLAST scores of searches with E(Z) and ESC proteins were much higher than the scores for searches with any of the other PcG proteins. In addition, two sequences with significant similarity to *E(Pc)* were detected. The two Arabidopsis proteins with significant similarity to E(PC) (AAG10815 and AAC17077) are much shorter than the fruit fly *E(Pc)*, which is 2,033 amino acids. Further analysis of these sequences indicates that they contain the domains found in *E(Pc)* homologs from fruit fly, mouse, *C. elegans*, and *Saccharomyces cerevisiae*.

The top two BLAST hits for all Polycomb proteins were analyzed based on length, organization of the gene, and similarity of conserved domains. Based on BLAST scores, gene organization, and conservation in critical domains, we determined that only homologs of *E(z)*, *esc*, and *E(Pc)* exist in Arabidopsis. Relatively strong BLAST hits to *pleiohomeotic* (*pho*), *mx*, *Pc*, and *p* were found, but subsequent analysis determined that they were not PcG homologs. The Arabidopsis proteins most similar to *pho* showed homology within a zinc finger domain, but not to any other regions of *pho*. Therefore, these zinc-finger proteins were determined to not be orthologs of *pho*. The proteins most similar to *mx* contained an RNA binding motif, but were different in length and composition throughout the rest of the protein and were clearly not orthologs of *mx*. The search for *Pc* homologs identified chromodomain-containing proteins that were members of other types of gene families. The lack of homology between *Pc* and these proteins outside the chromodomain clearly indicated that no *Pc* homologs exist in Arabidopsis. *P* candidates were found to have homology in a RING finger domain, but no homology outside this domain.

Table 1. *PcG homologs in the Arabidopsis genome*

A search for Polycomb-group homologs in Arabidopsis. The protein sequences of the fruit fly PcG proteins were used to perform TBLASTN searches of the Arabidopsis genomic sequence. The fruit fly sequences used were as follows: *esc*, S58672; *E(z)*, AAC46462; *E(Pc)*, AAF58641; *pho*, AAF59378; *mxc*, AAF27644; *Pc*, CAA39229; *psc*, CAA41965; *asx*, CAA04568; *pcl*, AAA64457; *ph*, CAA45211 and S23632; *scm*, AAB57632. The accession no. for the Arabidopsis sequence with the best BLAST score for each search is shown in the second column (the BLASTP score in parentheses). The number of Arabidopsis homologs for each PcG protein is indicated in the last column. The *esc* and *E(z)* homologs have been previously documented (Goodrich et al., 1997; Grossniklaus et al., 1998; Ohad et al., 1999). The accession nos. for the predicted protein sequences of the two *E(Pc)* homologs found in the Arabidopsis genome are shown in parentheses.

PcG Proteins	Best Score in Arabidopsis	Likely Arabidopsis Homologs
<i>esc</i> (extra sex combs)	AF129516 (4e-73)	<i>Fie</i>
<i>E(z)</i> (Enhancer of zeste)	AF100163 (7e-69)	<i>Clf</i> , <i>Eza1</i> , and <i>Medea</i>
<i>E(p)</i> enhancer of polycomb	AC011808 (2e-20)	AAG10815 and AAC17077
<i>pho</i> (pleiohomeotic)	AL391716 (4e-13)	0
<i>mxc</i> (multi sex combs)	T00677 (2e-10)	0
<i>Pc</i> (Polycomb)	AB006706 (5e-07)	0
<i>psc</i> (posterior sex combs)	AB026655 (2e-06)	0
<i>asx</i> (additional sex combs)	AC007651 (0.18)	0
<i>pcl</i> (polycomblike)	AL132966 (0.019)	0
<i>ph</i> (polyhomeotic)	AC006592 (0.23)	0
<i>scm</i> (sex combs on midleg)	AJ292982 (0.82)	0
super sex combs (<i>sxc</i>)	Not cloned in fruit fly	N.A.
sex combs extra (<i>sce</i>)	Not cloned in fruit fly	N.A.

Therefore, these candidates were determined to be RING finger proteins, but not *Psc* homologs.

The 11 cloned PcG proteins from fruit fly and the *E(z)*, *esc*, and *E(Pc)* homologs from Arabidopsis were used as queries to perform TBLASTN searches against the public maize expressed sequence tag (EST) database and the Pioneer Hi-Bred database. Maize homologs were found for *E(z)*, *esc*, and *E(Pc)*, but not for *Pc*, *ph*, *Psc*, *Polycomblike* (*Pcl*), *Additional Sex combs* (*Asx*), *Scm*, *pho*, or *mxc*. The full-length sequences of the maize *E(z)*, *esc*, and *E(Pc)* genes were then obtained and characterized.

Maize Has Three *E(z)* Homologs

The *Mez1* cDNA is 3,180 bp in length and produces a predicted protein of 933 amino acids (Fig. 1). The *Mez2* cDNA is 3,025 bp in length and encodes a putative protein of 893 amino acids. The *Mez3* cDNA is 3,149 bp in length and encodes a putative 896-amino acid protein. *Mez1* maps to the short arm of chromosome 6 (bin 6.01–6.02) and *Mez2* maps to the short arm of chromosome 9 (bin 9.04; data not shown). *Mez3* has not been mapped.

The *MEZ2* and *MEZ3* putative proteins are 89% identical and the nucleotide sequences of the genes are 92% identical. Based on the high degree of sequence homology, *Mez2* and *Mez3* are likely the genome duplicates resulting from the paleotetraploid origin of maize. It is common to find two closely related sequences in the maize genome due its evolutionary history (Gaut and Doebley, 1997). The amino acid sequences of *MEZ1* and *MEZ2* are 42% identical and 56% similar to each other over their entire lengths. The nucleotide sequences of *Mez1* and *Mez2* are 52% identical.

Arabidopsis contains at least three *E(z)*-like genes, and mutations in two of these genes, *Mea* and *Clf*, show distinct phenotypes. We attempted to determine which of the Arabidopsis *E(z)*-like genes that *Mez1*, *Mez2*, and *Mez3* were most closely related to. The SET domain sequences of all *E(z)*-like proteins (indicated in Fig. 1) were aligned using ClustalW. This alignment was then processed using PHYLIP, and a parsimonious tree was constructed (Fig. 2). The tree shows grouping of the Arabidopsis *clf* and the maize *Mez1*. When the full-length protein sequences were used for the alignments, the same tree was produced. This suggests that *Mez1* is a *clf*-like gene in maize, whereas *Mez2* and *Mez3* are likely to be *Eza1* homologs. The low degree of nucleotide similarity between *Mez1* and *Mez2/3* corroborates the notion that these genes may have distinct evolutionary origins. The orthology of monocot *E(z)* homologs with *Eza1* and *Clf1*, rather than *Medea* is supported by overall sequence similarity as well as by the fact that *Medea* has approximately 300 fewer amino acids between the EZD2 and SANT domains, which are present in the monocot proteins (Fig. 3). No gene more similar to *Medea* than to *Eza1* or *Clf1* was detected in any monocot EST or genomic sequence.

Alignment of *E(z)* Homologs Identifies Five Conserved Domains

Alignments of plant and animal *E(z)* homologs were used to identify conserved domains (Fig. 1). We first searched for previously annotated domains in the *MEZ* protein sequences using the SMART program (Schultz et al., 2000). A Cys-rich region and a SET domain (Fig. 3) near the C terminus of the proteins are conserved among all *E(z)* homologs. The

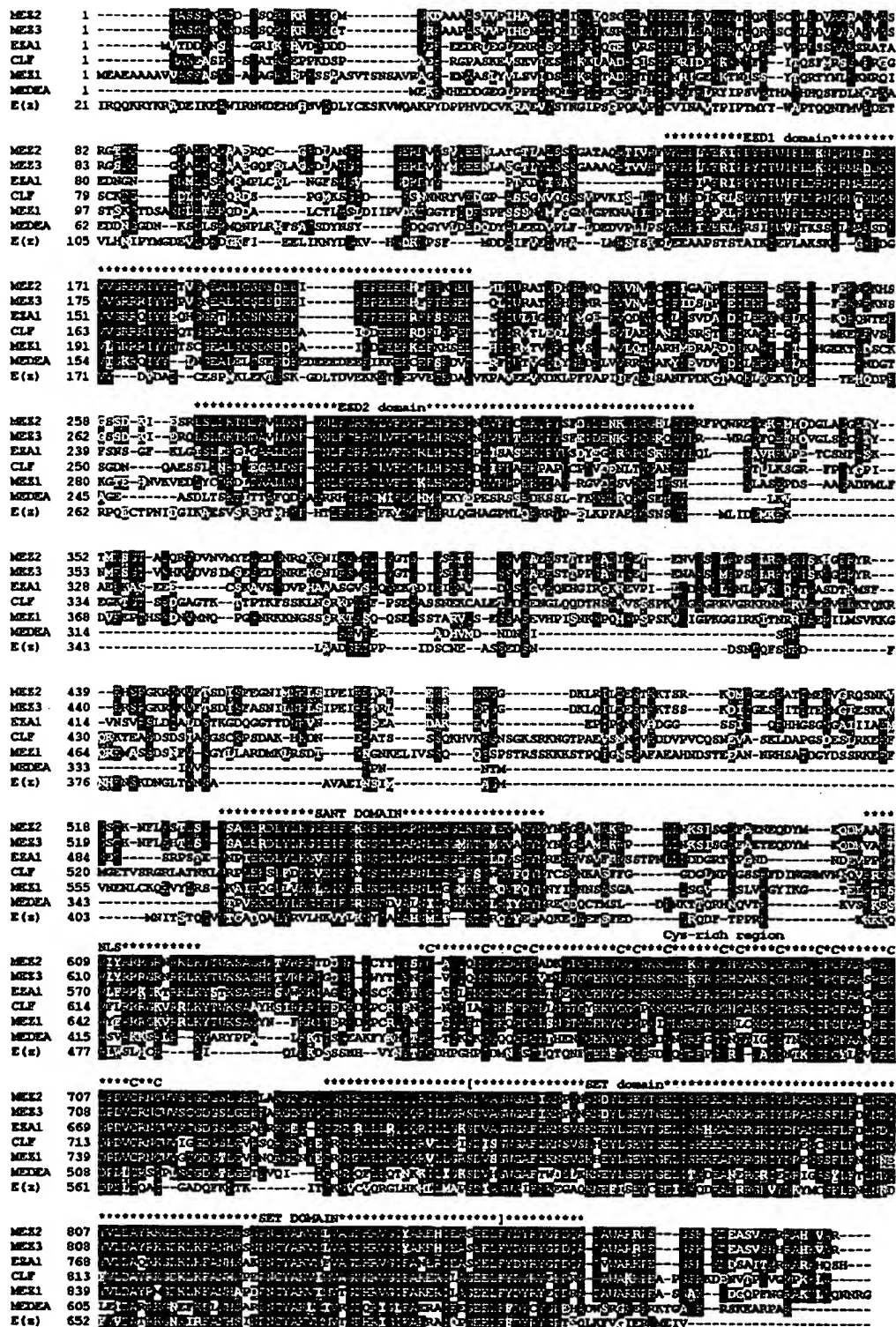


Figure 1. Alignment of plant and animal E(z)-like sequences identifies two new conserved domains, EKD1 and EKD2. The sequences of the maize E(z)-like proteins, MEZ1, MEZ2, and MEZ3, were aligned with the Arabidopsis E(z)-like proteins, CLF (AAC23781), MEDEA (AAC39446), and EZA1 (T01127), and the fruit fly E(z) sequence using ClustalW. The alignment was then colored using Boxshade (http://www.ch.embnet.org/software/BOX_form.html) to shade conserved residues in black and similar amino acids in gray. The position of the EKD1, EKD2, SANT Cys-rich, and SET domains are indicated above the alignments. A putative nuclear localization signal (NLS) is also indicated. The brackets identify the region for the phylogenetic analysis shown in Figure 2.

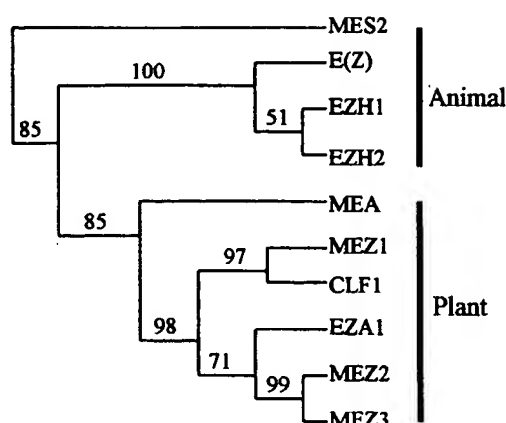


Figure 2. Relationships of maize and Arabidopsis *E(z)*-homologs. The SET domains (indicated in Fig. 1) from plant and animal homologs of *E(z)* were aligned using ClustalW and were analyzed using PHYLIP parsimonious methods to generate a parsimonious tree. The bootstrap values are indicated at nodes in the tree. The sequences used are MES-2 (AAC27124), EZH1 (AAC50778), EZH2 (AAC51520), *E(Z)* (AAC46462), MEZ1 (AF443596), MEZ2 (AF443597), MEZ3 (AF443598), CLF1 (AAC23781), EZA1 (T01127), and MEDEA (AAC39446). The phylogenetic analysis reveals that MEZ1 is most orthologous to CLF1, whereas MEZ2 and MEZ3 are orthologs of EZA1.

Cys-rich region has 15 invariant Cys residues with a conserved spacing pattern in all *E(z)* homologs. The spacing of the Cys residues in all *E(z)* homologs is unique and is different from other Cys-rich zinc finger domains involved in DNA binding. The SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain found at the C-terminal end of the protein is also highly conserved among all *E(z)* homologs. In addition to the SET domain, a SANT (SWI3, ADA2, N-CoR, and TFIIB' DNA-binding domains) domain was identified by SMART in the plant and animal proteins (Fig. 3). SANT domains are often involved in nonspecific DNA binding (Aasland et al., 1996).

In addition to the domains identified by SMART, two additional regions of conservation are present in the plant and animal *E(z)*-like proteins. These domains were analyzed by BLASTP and are not found in any other sequences outside of *E(z)* homologs. Therefore, we have named these two domains Enhancer of zeste domain1 (EZD1) and Enhancer of zeste domain2 (EZD2) because they represent conserved domains specific to the *E(z)* family. EZD1 is a highly conserved acidic region of 70 amino acids in the N-terminal region (Figs. 1 and 3), originally noted by Grossniklaus et al. (1998) as an acidic domain. The EZD1 domain contains a significant proportion of charged residues (34%–39%), with seven more acidic residues than basic residues. The function of this domain is not known. EZD1 is highly conserved between MEZ1, MEZ2, MEZ3, CLF, and EZA1. Although the primary sequence is less conserved in MEA and animal *E(z)*-like proteins, a similar distribution of charged residues exists. EZD2 is a small,

highly conserved region of 44 amino acids near amino acid 250 of the plant and animal *E(z)*-like proteins. This corresponds to the C₅ region noted by Goodrich et al. (1997). The EZD2 domain contains five highly conserved Cys residues and is composed primarily of polar or charged residues.

Maize Contains Two ESC/FIE Homologs

Two homologs of fruit fly *esc* were isolated from maize, *ZmFie1* (*Zea mays* fertilization independent endosperm 1) and *ZmFie2* (*Zea mays* fertilization independent endosperm 2). The *ZmFIE1* and *ZmFIE2* proteins are 76% identical and 85% similar over their entire lengths (Fig. 4). The nucleotide sequences of *ZmFie1* and *ZmFie2* are 83% identical to one another. *ZmFie1* maps to chromosome 4 (bin 4.05), and *ZmFie2* maps to chromosome 10 (bin 10.03; data not shown). It is likely that *ZmFie1* and *ZmFie2* are genome duplicates that arose from the ancient polyploidization of maize. The map position of these genes is consistent with ancient polyploidy as the origin of the duplication (Gaut, 2001).

An alignment of *esc*-like proteins is shown in Figure 4A. All previously cloned *esc*-like proteins contain seven WD-40 repeats (Ng et al., 1997; Korf et al., 1998; Ohad et al., 1999). *ZmFIE1* and *ZmFIE2* also contain seven WD-40 repeats, which are indicated in Figure 4. The WD-40 repeats are more highly conserved between the maize *ZmFIE* proteins and the fruit fly *ESC* and Arabidopsis *FIE* than other regions of the proteins. A putative nuclear localization signal is also found near the N terminus of the plant *esc*-like proteins. The *ZmFIE1* protein contains a unique 58-amino acid extension at the N-terminal end and a unique 22-amino acid extension at the 3' end relative to *ZmFIE2*.

Plants Contain Homologs of *E(Pc)*

Two predicted proteins with significant similarity to *E(Pc)* were detected in the Arabidopsis genome.

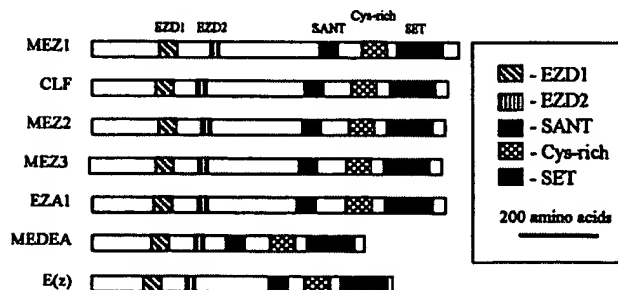
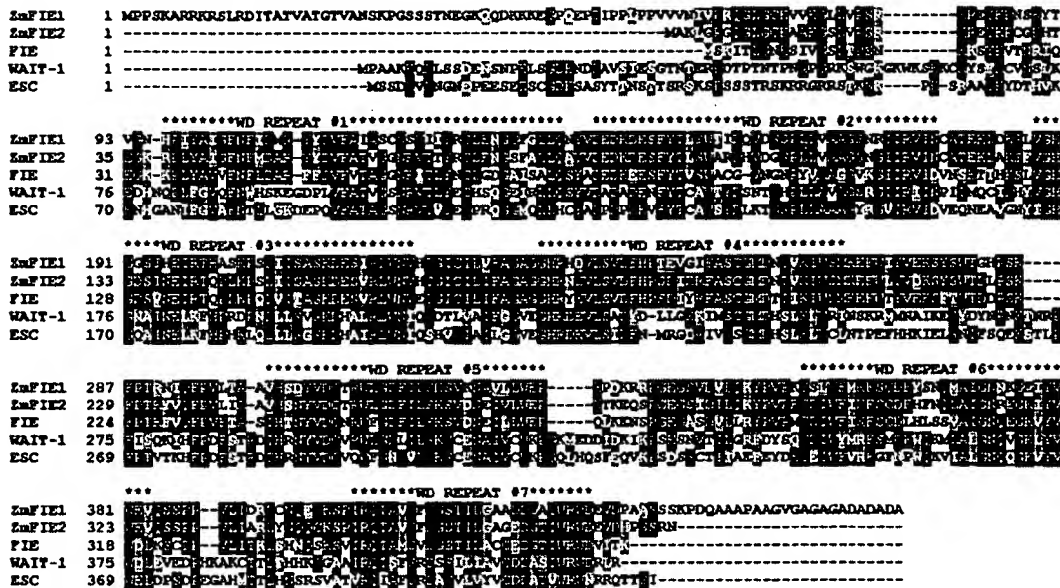


Figure 3. Schematic diagram of *E(z)* homologous proteins. The relative positions of the EZD1, EZD2, SANT, Cys-rich, and SET domains found in MEZ1, MEZ2, CLF1, MEDEA, EZA1, and *E(Z)* are indicated by the shaded boxes. Identities of each protein are shown at the right, and the structures are oriented with the amino terminal on the left and the carboxy terminal on the right. The legend at the right indicates the shading pattern for each domain and the scale of the drawings.

A. Alignment of esc-like sequences



B. Schematic diagram of ESC proteins

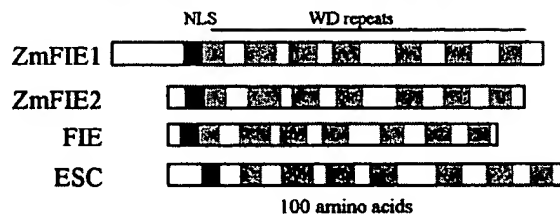


Figure 4. Maize contains two esc homologs. A, The protein sequences of the two maize esc homologs, ZmFIE1 (AY061964), ZmFIE2 (AY061965), Arabidopsis FIE (AAD23584), fruit fly ESC (AAF53124), and the human WAIT-1 (AAC68675) were aligned using ClustalW. The identities and amino acid position for each sequence is shown on the left. The alignment was shaded with Boxshade such that the conserved residues are colored black and the similar residues are gray. The location of the WD40 repeats is indicated above the alignment. B, The protein sequences of ZmFIE1, ZmFIE2, FIE, ESC, and WAIT-1 are shown as schematic diagrams with the location of the WD40 repeats and a putative NLS indicated. The position of a putative NLS is indicated by a black box, and the gray boxes indicate the positions of the WD repeats. The scale of the drawings is indicated.

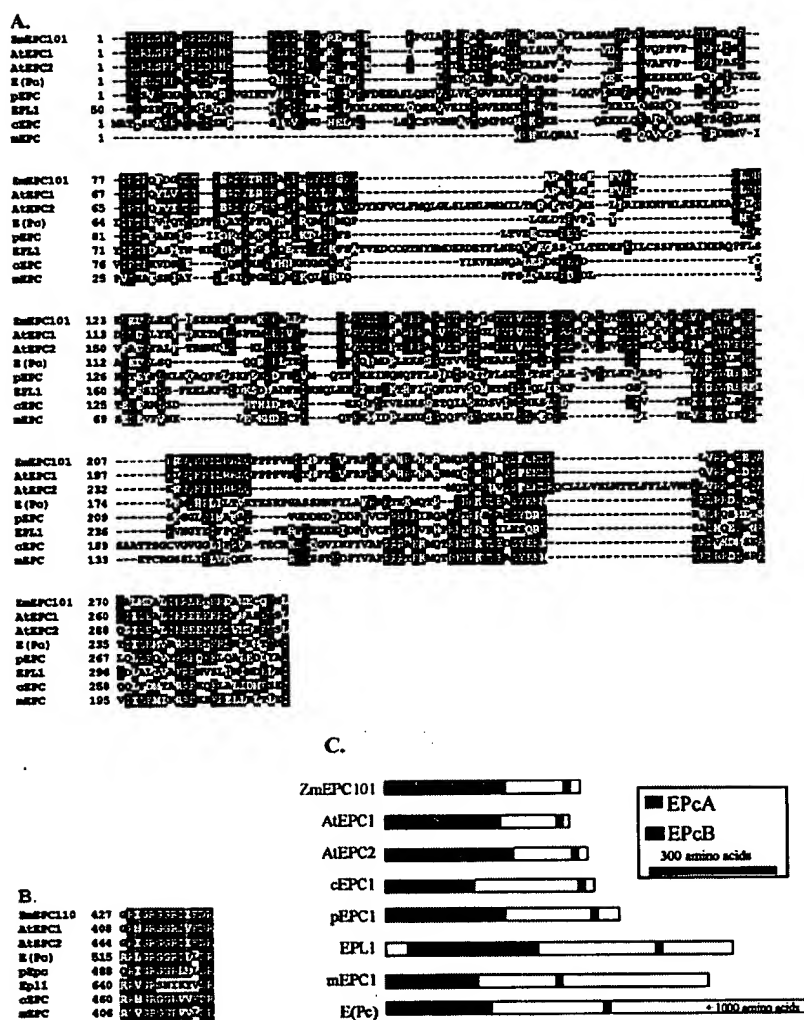
These proteins were named AtEPL1 (AAG10815) and AtEPL2 (AAC17077). *AtEpl1* is located on chromosome 1 at 25cM (BAC F19K19) and is represented by a single EST (AV559019) from green silique tissue. *AtEpl2* is also located on chromosome 1 at 117 cM (yeast artificial chromosome YUP8H12R) and is not represented by any Arabidopsis ESTs. The Arabidopsis proteins are 67% identical and 79% similar to each other.

We searched for ESTs from maize with significant similarity to the Arabidopsis Epc genes. A single EST (BE511872) with significant similarity to the *AtEpl1* and *AtEpl2* genes was found. The maize *E(Pc)* homolog, *ZmEpl101*, is 52% identical and 70% similar to *AtEpl2*. In addition to the *ZmEpl101* EST from maize, *E(pc)* homologous ESTs were also found in soybean

(*Glycine max*), tomato (*Lycopersicon esculentum*), potato (*Solanum tuberosum*), wheat (*Triticum aestivum*), and barley (*Hordeum vulgare*). This indicates that *E(Pc)* homologs are present and expressed in a variety of plant species.

E(Pc) homologs have previously been identified in mammals, yeast, and *C. elegans* (Stankunas et al., 1998). Alignments of the known *E(Pc)* homologs with the plant *EPL* proteins revealed significant conservation of the previously characterized EPcA domain (Stankunas et al., 1998; Fig. 5A). A portion of the EPcB domain is strongly conserved in all eukaryotes with the exception of the yeast *Epl1* gene (Fig. 5B). The EPcC domain identified in fruit fly and mammals (Stankunas et al., 1998) is not observed in yeast, *C. elegans*, or higher plants.

Figure 5. Plants contain sequences similar to *E(Pc)*. A, The protein sequences the EpcA domain of ZmEPL101 (AF443599), Arabidopsis EPL1 (AtEPL1-AC011808), Arabidopsis EPL2 (AtEPL2-YUP8H12R), *C. elegans* EPC1 (cEPC1-CAC35840), *Schizosaccharomyces pombe* EPC1 (pEPC1-T41631), *S. cerevisiae* EPL1 (NP 116629), *M. musculus* EPC1 (mEPC1-AAC64272), and fruit fly *E(Pc)* (AAF58641) were aligned using ClustalW. The identity and position within the protein for each sequence is shown at the left. This alignment was shaded using Boxshade to show conserved residues in black and similar residues in gray. B, The portion of the EpcB domain that is conserved in plants and animals was aligned using ClustalW. The protein and amino acid position in the protein are indicated on the left. C, The structures of the EPL proteins are represented by drawings. The size and location of the EpcA and EpcB domains are indicated by the shaded boxes. The identity of each structure is shown on the left and the structures are orientated with the amino terminus on the left. Only the N-terminal 1,023 amino acids of the fruit fly *E(Pc)* are represented. The legend on the right indicates the shading used for each domain and the scale of the drawings.



Mez Genes in Maize Are Expressed throughout Plant Development, and *Mez2* Shows a Tissue-Specific Alternative Splicing Pattern

The expression level of the maize PcG homologs is low and difficult to detect by northern-blot analysis. In addition, it is not possible to generate specific probes for the pairs of duplicate genes, *ZmFie1/ZmFie2* and *Mez2/Mez3*, due to the high degree of nucleotide identity between these two sequences. Therefore, we used reverse transcriptase (RT)-PCR analysis to determine the expression pattern of the PcG genes of maize (Fig. 6). *ZmEpl101*, *Mez1*, *Mez2*, and *Mez3* transcripts were successfully amplified in all tissues tested. In most cases, there was some variation in the amount of product amplified from different tissues; this may reflect slight differences in the amount of cDNA used in the amplification reaction and not the exact expression level.

Multiple transcripts were observed when the entire coding sequence of *Mez2* was PCR amplified (data not shown). Three transcripts were characterized from *Mez2*, including the full-length transcript

and two alternatively spliced products (Fig. 7A). The alternative splice products are named *Mez2 alternative splice 1* (*Mez2^{as1}*), which is 2,385 bp in length and *Mez2 alternative splice 2* (*Mez2^{as2}*), which is 2,234 bp in length. Relative to *Mez2*, *Mez2^{as1}* lacks a 659-bp fragment and *Mez2^{as2}* is missing a fragment of 810 bp. The missing fragment in *Mez2^{as1}* corresponds to basepair 1,017 to 1,675 of the full-length transcript and causes a frameshift resulting in the production of a truncated protein of 341 amino acids (Fig. 7B). The deletion in the *Mez2^{as2}* corresponds to basepair 1,017 to 1,826 of the full-length transcript and results in a 624-amino acid protein that is missing the SANT domain (Fig. 7B).

We tested cDNA from eight different maize tissues to determine whether the splicing of *Mez2* transcripts is tissue specific (Fig. 7A). Amplification from whole kernel (11 d after pollination [DAP]), 10-d whole seedling, and immature leaf cDNA revealed the presence of the three transcripts observed previously. Low levels of *Mez2^{as1}* and *Mez2^{as2}* transcripts were detected in endosperm

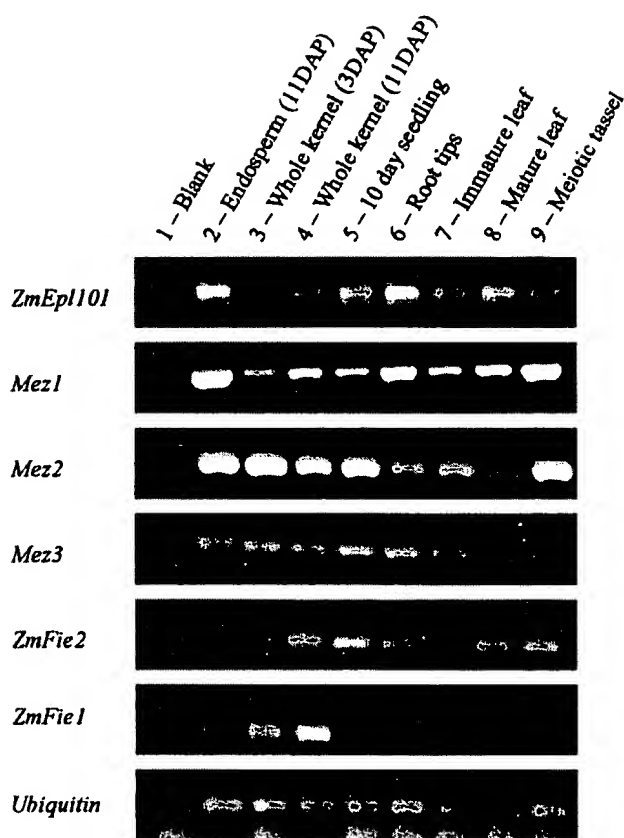


Figure 6. Expression pattern of the maize PcG genes. RT-PCR was performed on RNA isolated from various maize tissues to determine the expression patterns of the maize PcG genes. The sequence amplified is indicated on the left, and the source of the RNA for each lane is indicated above the pictures (1, blank; 2, endosperm (11 DAP); 3, whole kernel (3 DAP); 4, whole kernel (11 DAP); 5, 10-d-old seedling; 6, root tips; 7, immature leaf; 8, mature leaf; and 9, meiotic tassel).

cDNA (11 DAP), whereas no band corresponding to the full-length transcript was observed. We cannot rule out the presence of a low level of the full-length transcript in this tissue. *Mez2^{as1}* and *Mez2^{as2}* transcripts were not observed in whole kernel tissues isolated 3 DAP. In root tips, mature leaf, and meiotic tassel, the *Mez2^{as1}* was not observed. The presence of the same alternative splice products has been noticed in multiple genotypes (B73, Mo17, and W22) and in tissue taken from different sources (data not shown). Although it is possible that the alternative splice pattern is influenced by environmental stimuli, we have not noticed differences in the splicing patterns taken from the same tissues (data not shown).

***ZmFie2* Is Expressed throughout Development, Whereas *ZmFie1* Expression Is Limited to the Early Embryo and Endosperm**

The presence of *ZmFie1* and *ZmFie2* transcripts in various plant tissues was tested by RT-PCR (Fig. 6).

ZmFie2 transcripts were detected in all tissues tested. This is similar to the expression pattern of the *Mez* genes. *ZmFie1* transcripts were only found in kernels after pollination, but not in any other plant tissues tested. The *ZmFie1* transcripts were detected in embryo (data not shown) and endosperm tissue.

DISCUSSION

PcG proteins play an important role in the maintenance of gene silencing in animals. In this study, we searched the complete Arabidopsis genome sequence and maize EST databases for homologs of PcG proteins. This exhaustive search identified homologs for only three of the 11 cloned fruit fly PcG proteins, *E(z)*, *esc*, and *E(Pc)*. Biochemical studies have found that the PcG proteins of animals are organized into at least two complexes, the *E(z)/esc* complex (Jones et al., 1998; Sewalt et al., 1998) and the PRC1 complex, which includes *Pc*, *Ph*, *Scm*, and *Psc* (Shao et al., 1999). The role of the E(PC) protein is not well characterized, and current evidence suggests that it operates independently of the E(Z)/ESC complex. Genetic evidence from fruit fly suggests that the

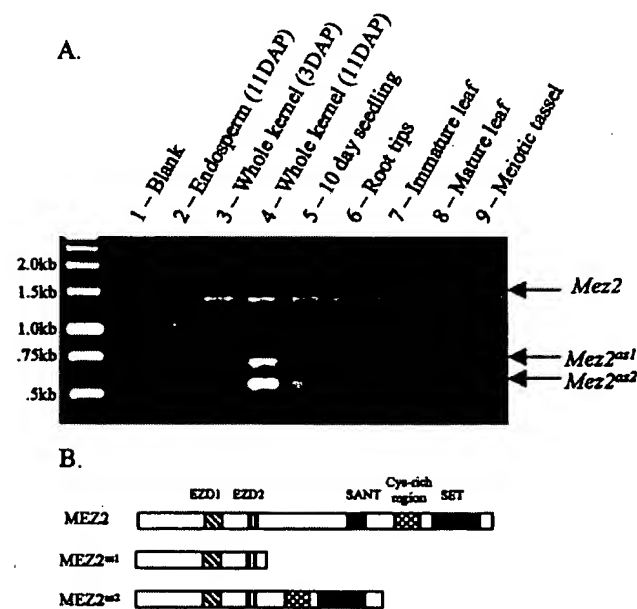


Figure 7. Alternative splicing of *Mez2*. A, PCR amplification of *Mez2* transcripts from various maize tissues. The expected size of the amplified band is 2.2 kb. The presence of multiple bands indicates alternative splicing of the *Mez2* transcript. Different tissues displayed different patterns of splicing for *Mez2*. The three bands amplified from ear tissue (indicated by the arrows at the right of the picture) were sequenced and are designated as *Mez2*, *Mez2^{as1}*, and *Mez2^{as2}*. The products amplified from embryo tissue were also cloned and sequenced and were all found to correspond to *Mez2^{as2}*. B, Schematic diagrams of the alternative splicing event and the altered proteins. The shading pattern is identical to that used in Figure 3. *Mez2^{as1}* introduces a frameshift and will produce a truncated protein that lacks the SANT, NLS, Cys-rich, and SET domains. *Mez2^{as2}* produces an in-frame deletion that removes only the SANT domain.

functions of all PcG proteins are necessary for repression of homeotic genes (Simon, 1995). The genetic data include similar phenotypes of single mutants as well as studies of double mutants. The observations that *E(z)* and *esc* homologs function to repress gene expression in organisms that lack other PcG proteins, such as plants and *C. elegans* (Goodrich et al., 1997; Holdeman et al., 1998; Kelly and Fire, 1998; Korf et al., 1998), indicates that the *E(z)/esc* complex is capable of repression of transcription in the absence of the PRC1 complex.

In our searches, we failed to find homologs of any proteins in the PRC1 complex. The fact that the *E(Z)/ESC* complex is capable of repressing transcription in the absence of PRC1 raises the question of what, if anything, is fulfilling the function of the PRC1 complex in plants and *C. elegans*. No catalytic activities have been defined for any of the fruit fly proteins found in the PRC1 complex. In vitro, the PRC1 complex of fruit fly interferes with SWI/SNF-dependent chromatin remodeling activities (Shao et al., 1999). If the primary role of the PRC1 complex is inhibition of SWI/SNF-dependent chromatin remodeling, then plants may contain a different group of proteins, targeted by *E(Z)/ESC*, that interfere with SWI/SNF activity. In an alternate manner, the *E(Z)/ESC* complex may be capable of silencing transcription independent of PRC1 function. The role of the PRC1 complex in fruit fly and mammals may be to enhance or stabilize this silencing, whereas in plants and *C. elegans* this supplementary function is absent.

The *E(Z)* and ESC Proteins Can Provide Two Mechanisms for Repression

We have found multiple homologs of *E(z)* and *esc* in maize, and previous studies have documented similar homologs in Arabidopsis (Goodrich et al., 1997; Grossniklaus et al., 1998; Ohad et al., 1999; Preuss, 1999). Our sequence analysis of the MEZ1, MEZ2, and MEZ3 proteins revealed the presence of multiple domains that are conserved between plant and animal *E(Z)*-like proteins. The most conserved domain is the SET domain, which is found near the C terminus of all *E(Z)*-like proteins. The SET domain of the *Su(var)3-9*-like proteins and mammalian G9a protein has been demonstrated to act as a histone methyltransferase (Rea et al., 2000; Lachner et al., 2001; Nakayama et al., 2001; Tachibana et al., 2001). These proteins contain Cys-rich regions on both sides of the SET domain. The presence of both Cys-rich regions was proposed to be required for targeting histone methyltransferase activity to histone H3 (Rea et al., 2000). In this same study, other proteins such as EZH2 and HRX, which contain only one Cys-rich region, did not possess detectable histone methyltransferase activity. The maize *E(Z)*-like proteins like fruit fly and mammalian *E(Z)* proteins only contain a Cys-rich region on the N-terminal side of the SET

domain. Due to the high conservation of the SET domain, it is likely that the plant *E(Z)*-like proteins are protein methyltransferases. Further research will determine if *E(Z)*-like proteins are capable of histone methylation in vivo or if they methylate other proteins.

Alignments of the two ZmFIE proteins with Arabidopsis FIE, fruit fly ESC, and mammalian WAT-1 revealed several conserved features. The maize ZmFIE1 and ZmFIE2 proteins contain seven WD repeats, the same as found in all other characterized ESC-like proteins (Ng et al., 1997; Ohad et al., 1999). The spacing of the WD repeats found in ZmFIE1 and ZmFIE2 is also consistent with that observed in other ESC-like proteins. This suggests that the ZmFIE proteins found in plants are likely to form a β -propeller structure similar to that predicted for fruit fly ESC (Ng et al., 1997). The sequence conservation of ZmFIE proteins with the fruit fly and mammalian proteins makes it reasonable to propose that these proteins are involved in similar protein-protein interactions. The ESC-like proteins have been documented to physically interact with two other proteins. A direct interaction between *E(Z)* and ESC homologs has been demonstrated in fruit fly, mice, *C. elegans*, and Arabidopsis (Jones et al., 1998; Sewalt et al., 1998; Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000; Xu et al., 2001). The ESC-like proteins in fruit fly and mouse directly interact with Rpd3 histone deacetylase proteins (van der Vlag and Otte, 1999; Tie et al., 2001). These interactions provide a mechanism for ESC-like proteins to link the catalytic activities of protein methylation by *E(Z)*-like proteins with histone deacetylation by Rpd3-like proteins.

The *E(Z)/ESC* complex could perform two distinct functions in repressing gene expression. First, the *E(Z)/ESC* contains a histone deacetylase protein, which can mediate a repressive chromatin structure. Histone acetylation states are relatively unstable and require constant presence of a histone deacetylase to be maintained (Jenuwein, 2001). Second, the SET domain of *E(z)* is predicted to be involved in protein methylation. Unlike acetylation, methylation often tends to be more stable (Jenuwein, 2001). Although the role of histone acetylation in regulating chromatin states is defined (Cheung et al., 2000), the role of *E(z)*-mediated protein methylation is less apparent. Protein methylation may be involved in potentiating interactions with other proteins by creating specific binding sites. For example, the SET domain protein *Su(var)3-9* methylates Lys 9 of histone H3 (Rea et al., 2000). Heterochromatin protein HP1 will bind specifically to methylated histone H3, but not to unmodified histone H3 (Bannister et al., 2001; Lachner et al., 2001; Nakayama et al., 2001). In animals, the PcG proteins include a SET domain protein, *E(z)*, and a chromodomain protein, Polycomb. Homologs of the *Polycomb* gene are not present in plants, but other chromodomain containing genes, such as chromom-

ethylases, are present. Therefore, the E(Z)/ESC complex provides a combination of a reduction in histone acetylation coupled with targeted protein methylation that likely results in a stable repressive chromatin state.

Plants Contain Proteins Similar to E(Pc)

We have identified homologs of a third PcG protein, *E(Pc)*, from Arabidopsis and maize. The Arabidopsis *Epl1* sequences that we found are located at two unlinked locations on chromosome 1. The maize *ZmEpl101* gene is homologous to the Arabidopsis genes throughout the coding sequence. Unlike *E(z)* and *esc*, the size and organization of *E(Pc)* homologs differs between species. The fruit fly E(PC) protein is 2,033 amino acids. The homologs of *E(Pc)* from other species lack the C-terminal 1,400 amino acids present in the fruit fly protein and average 560 amino acids in length (Stankunas et al., 1998). All of the *E(Pc)* homologs contain two conserved domains, EpcA and EpcB. The EpcA domain encompasses approximately the first 200 amino acids of all *E(Pc)* homologs, whereas the EpcB domain is 13 amino acids. Stankunas et al. (1998) identified the EPcC domain as a region of conservation between the fruit fly and mammalian E(PC) protein sequences that is not present in *C. elegans* or *S. cerevisiae* homologs. Our alignments show that like the *C. elegans* or *S. cerevisiae* homologs, the plant *E(Pc)* homologs do not contain the EPcC domain identified in fruit fly and mammalian sequences (Stankunas et al., 1998).

To date, E(PC) has not been found in complexes associated with any other PcG proteins. Unlike the other PcG group genes, mutations in *E(Pc)* do not display homeotic transformations alone (Sato et al., 1983). However, mutations in *E(Pc)* enhance the homeotic transformations observed when other PcG genes are mutated (Sato et al., 1984). This suggests that *E(Pc)* plays a role distinct from other PcG proteins. In addition, a homolog of *E(Pc)*, *Epl1*, has been found in *S. cerevisiae*, a species that does not contain homologs of any other PcG proteins (Stankunas et al., 1998). The presence of an *E(Pc)* homolog in a species that lacks all other PcG proteins suggests that *E(Pc)* homologs may have a basic chromatin function independent of PcG proteins, but that is required for PcG silencing. In support of the idea that *E(Pc)* plays a broader role in chromatin regulation, the human *E(Pc)* homolog has been shown to repress and activate transcription (Shimono et al., 2000). In a similar manner, another study found that the yeast *E(Pc)* homolog *Epl1* was present in the NuA4 histone acetyltransferase complex involved in transcriptional activation (Galarneau et al., 2000). The plant *E(Pc)* homologs *Epl1*, *Epl2*, and *ZmEpl101* are likely to be proteins involved in chromatin-based regulation of gene expression, but may operate independently of *E(z)* and *esc* homologs.

Expression of Multiple E(Z) and ESC Proteins May Allow Specialization of Function in Plants

Plants contain multiple *E(z)* and *esc* homologs. Arabidopsis contains three homologs of *E(z)*: *Clf*, *Eza1*, and *Mea*. We have documented the presence of three *E(z)* homologs in maize. The first gene, *Mez1*, does not have a closely related nucleotide sequence in the maize genome and is most likely to be the maize ortholog of *Clf* based on our phylogenetic analysis. The other two *E(z)* homologs, *Mez2* and *Mez3*, are highly related to each other and are predicted to be the orthologs of *Eza1* based on sequence phylogeny. No *Mea* homologs are observed in the public monocot EST databases, or rice (*Oryza sativa*) genomic sequences (searches done February 8, 2002). This could simply reflect the fact that *Mea* is not highly expressed and is therefore not represented as an EST, or it could indicate that monocots do not have a *Mea* homolog. If there is no *Mea* present in monocots, another *E(z)* protein is likely performing a function in monocots analogous to *Mea*'s function in dicots. This could also reflect differences in the regulation of development in monocots and dicots.

In addition to having three distinct genes encoding *E(z)* homologs, maize also displays alternative splicing of at least one *E(z)* gene, *Mez2*. The alternative splicing pattern of *Mez2* could provide distinct functions arising from the same gene in different tissues and cells. We successfully amplified transcripts corresponding to multiple splice products from many tissues. However, the alternative splicing pattern was distinct in some tissues, indicating a developmental control of the splice pattern of *Mez2*. The presence of multiple transcripts in some tissues may be due to simultaneous presence of these transcripts in all cells of that tissue or may be due to cell type-specific expression of certain transcripts of *Mez2*.

The putative proteins produced by the alternatively spliced transcripts of *Mez2* are likely to provide different functions. The MEZ2^{as2} protein is similar in size and domain composition to MEA (with the exception of the SANT domain). The MEZ2^{as2} form predominates in embryonic tissues, which is the location of *Mea* function in Arabidopsis. It is possible that through alternative splicing, *Mez2* provides *Mea* and *Eza1* functions in maize. The MEZ2^{as1} protein, which is present in kernel, seedling, and immature leaf tissue, lacks the C-terminal two-thirds of the protein, including the SANT, Cys-rich, and SET domains. The truncated protein produced from *Mez2*^{as1} transcripts might negatively regulate function of the *E(z)/esc* complexes by binding ESC proteins in a non-functional complex. The production of the MEZ2^{as2} protein, which is missing a 270-amino acid region that includes the SANT DNA-binding domain, but still possesses a SET domain, may alter the localization or activities of the complex. The alternative splic-

ing of *Mez2* may allow for this gene to provide distinct functions in different tissues.

In this study, we have also documented that maize contains multiple *esc* homologs, *ZmFie1* and *ZmFie2*. In the other organisms in which the PcG proteins have been studied (fruit fly, humans, mice, *C. elegans*, and *Arabidopsis*), only one *esc* homolog has been found. Based on the chromosomal locations and high degree of nucleotide identity, it is likely that the duplication of these genes is due to the allotetraploid history of maize (Helentjaris, 1995; Gaut and Doebley, 1997). This duplication event may have allowed specialization of the function of *ZmFie1* and *ZmFie2* in expression pattern or function. We have found that *ZmFie1* is only expressed in kernel tissues, whereas *ZmFie2* is expressed in all tissues tested. The sequence of *ZmFie1* is also different from *esc* homologs, containing N- and C-terminal extensions. *ZmFIE1* may have evolved kernel-specific functions that are important in the development of the maize endosperm or embryo.

The E(Z) and ESC proteins function as a complex. The domain of the E(Z) and MEA proteins that interacts with ESC and FIE has been mapped to the N-terminal region of the protein (Jones et al., 1998; Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000). All three of the MEZ proteins from maize as well as the two putative proteins produced by alternative splicing of *Mez2* contain the N-terminal region of the protein and therefore are predicted to interact with the maize *ZmFIE* proteins. Evidence from *Arabidopsis* indicates that MEA and EZA1 interact with FIE (Luo et al., 2000). Assuming that the three full-length MEZ proteins and the two alternative splice isoforms can interact equally with both *ZmFIE* proteins, up to 10 distinct E(z)/*esc* complexes could be formed. The multiplicity of complexes could allow for a number of specialized roles in regulating gene expression. Furthermore, the expression and splicing pattern of *Mez* and *ZmFie* genes is tissue specific, which would regulate the exact set of E(Z)/ESC complexes present in a given tissue.

Epigenetic regulation of gene expression involving homologs of E(z) and *esc* is conserved across diverse species. These proteins offer a pathway for epigenetic regulation separate from DNA methylation. The main role of PcG proteins in plants and in animals may be to maintain the gene expression patterns determined by developmental decisions. This type of repression must be reset at meiosis each generation. In contrast, repression mediated by DNA methylation provides a meiotically heritable mechanism for gene silencing. There may be instances where PcG-mediated and DNA methylation-mediated silencing overlap, such as the regulation of AG and AP3 (Finnegan et al., 1996; Goodrich et al., 1997). However, we believe PcG-mediated silencing is primarily a source of epigenetic memory during development, whereas DNA methylation plays a

central role in heritable gene silencing for genome protection.

MATERIALS AND METHODS

Cloning of *Mez* Genes

Fruit fly (*Drosophila melanogaster*) E(z) (AAC46462) was used in a TBLASTN search of the Pioneer Hi-Bred EST database. Three contigs with significant similarity to E(z) were discovered and named *Mez1*, *Mez2*, and *Mez3*. Other SET domain-containing proteins were also identified, but were not included in this research because they had greater similarity to proteins other than to E(z). RACE was performed to obtain full-length cDNA sequence of *Mez1* and *Mez2*. Full-length *Mez3* was obtained from an EST clone that contained the entire coding region. RACE reactions were performed using the Marathon cDNA kit (CLONTECH, Palo Alto, CA) on cDNA produced from 1-week-old Mo17 seedlings. Advantage2 polymerase (CLONTECH) was used in the RACE reactions. The primers used in the RACE reactions were *Mez1F1* (5'-GGGTGTGGTGATGGTACATIGG-3'), *Mez1R2* (5'-CAGCTTGTCACCCATTCTGTATGCG-3'), *Mez2R3* (5'-TGCCTCGTCCTTCTTTGATCCTTCG-3'), and *Mez2F3* (5'-CTCACAAGGAAGCAGACAAACGCGG-3'). RACE products were gel purified and cloned into pGEM-T Easy (Promega, Madison, WI). All sequencing was performed using BigDye terminator cycle sequencing on an ABI sequencer (PerkinElmer Applied Biosystems, Foster City, CA). Sequencing reactions were done in a 10 μ L volume with 200 to 400 ng of DNA and 10 pg of primer. The cycling conditions used were 95°C for 2 min, 70 cycles at 95°C for 15 s, 55°C for 20 s, 60°C for 4 min, followed by 72°C for 7 min.

Cloning of *ZmFie* Genes

Arabidopsis FIE (AF129516) was used in a TBLASTN search of the Pioneer Hi-Bred maize (*Zea mays*) EST database. Two contigs with significant similarity were found and were named *ZmFie1* and *ZmFie2*. Both contigs contained at least one full-length EST. The sequence of the *ZmFie* genes was obtained by sequencing the full-length clones as described above.

Cloning of *ZmEpl101*

The *Arabidopsis* genome sequence was searched using fruit fly E(Pc) as a query. Two sequences were identified in this search, *AtEpl1* and *AtEpl2*. The AtEPL1 (AC011808) protein sequence was then used to perform a TBLASTN search of the public maize EST database (<http://www.zmdb.iastate.edu/>). One EST (BE511872) with significant similarity was found. We obtained the remaining sequence for the *ZmEpl101* gene by RACE-PCR. The primers used were *Mepc1R1* (5'-GTCCGGAGAAGAGGATTCCATCGATC-3') and *Mepc1R2* (5'-CCTCGTCCAATCCTACCTCGACACC-3').

Phylogenetic Analysis

The SET domains from all E(z)-like proteins were aligned using ClustalW (the exact region used is indicated

in Fig. 1). This alignment was then submitted to the PHYLIP server at <http://bioweb.pasteur.fr/sequal/phylogeny/phytip-uk.html>. The protpars feature was used, with bootstrapping performed before analysis. One hundred replicates were examined to determine bootstrap values. The consensus tree was then displayed with bootstrap values.

RT-PCR Analysis

RT-PCR was used to assess expression patterns due to the relatively low expression of the maize PcG homologs, and to the fact that most of the genes were duplicated. Total RNA was extracted with Trizol (Invitrogen, Carlsbad, CA) from 10 tissues from the inbred B73 (endosperm [11 DAP], whole kernel [3 DAP], whole kernel [11 DAP], 10 d seedling [whole plant included], root tips, immature leaf [leaves 3–5], mature leaf [fully expanded leaf 10], and meiotic tassel). One microgram of total RNA was used to make cDNA with the SMART cDNA synthesis kit according to manufacturer's instructions (CLONTECH). PCR reactions were performed in a 25- μ L total volume containing approximately 0.5 ng of cDNA, 5 pM each primer, 1 unit of *Taq* polymerase (Promega), 2.5 μ L of 10 \times reaction buffer, 2 μ L of 25 mM $MgCl_2$, and 0.3 μ L of 25 mM dNTPs. Primers used for the RT-PCR reactions were *Ubi1F1* (5'-TAA-GCTGCCGATGTGCCTGCGTCG-3') and *Ubi1R1* (5'-CT-GAAAGACAGCACATAATGAGCACAGGC-3') for *Ubiq-uitin*; *Mepc1F1* (5'-CCGGCCATGGCGAAGCTG-3') and *Mepc1R2* (5'-CCTCGTCCAATCCTACCTCGACACC-3') for *ZmEpl101*; *Mez1F1* (5'-GGGTGTGGTGATGGTACATTGG-3') and *Mez1R1* (5'-CGGGACCTAACTCTACGGA-TGG-3') for *Mez1*; *Mez2F8* (5'-CCCCTGTTTTGCAGCC-AGTCGTGA-3') and *Mez2R8* (5'-GGTGAGAGAAGGA-TGCCTCGTCC-3') for *Mez2*; *Mez3F3* (5'-AGTATGTG-TTGGATGCTTATCGCAAGG-3') and *Mez3R2* (5'-GGT-TGTCAGTTTGTACCTTCCGACC-3') for *Mez3*; *Mesc1-5* (5'-TTTGCAAGTTGTGGCATGGA-3') and *Mesc1R2* (5'-CCCAACTTTCAACATTCGAAGCATTTC-3') for *ZmFie2*; and *FLMesc2F1* (5'-CAACATCTGGCACAGCATGC-3') and *Mesc2R3* (5'-GTTGCCTATTGCCATCTGGTTGGAG-3') for *ZmFie1*. Conditions of the PCR were as follows: 94°C for 2 min, 35 cycles at 94°C for 30 s, 63°C for 30 s, and 72°C for 2 min, followed by 72°C for 7 min. Amplified products were separated in a 1% (w/v) agarose Tris borate-EDTA gel and were visualized by ethidium bromide staining. The primers chosen flanked introns and are expected to produce different sized transcripts from genomic DNA and cDNA. The *ZmEpl101* primers amplify a 450-bp cDNA product and an approximately 1,200-bp genomic product. The *Mez1* primers would amplify a 717-bp cDNA product and a 1,235-bp genomic product, the *Mez2* primers would amplify a 556-bp cDNA product and a 1,083-bp genomic product, the *Mez3* primers would amplify a 509-bp cDNA product and a 1,211-bp genomic product, the *ZmFie1* primers would amplify 1,135-bp cDNA product and a 4,008-bp genomic product, and the *ZmFie2* primers would amplify 711-bp cDNA product and a 1,748-bp genomic product.

Analysis of Mez2 Alternative Splicing

Alternative splicing of *Mez2* was indicated by the presence of multiple bands when using primers that amplified the entire coding sequence. The primers used were *Mez2F6* (5'-CGCAGCTGATACGGCAAGTCCAATCG-3') and *Mez2R2* (5'-GTATCATCCGGAGCGACTCTTCAGC-3'). These primers are expected to produce a 2,594-bp cDNA product and a >9-kb genomic fragment. To characterize the alternative splice products, PCR was performed using 10 ng of B73 cDNA isolated from ear tissue and Amplitaq Gold DNA polymerase (PerkinElmer Applied Biosystems). The cycling conditions were 95°C for 5 min, 35 cycles at 95°C for 30 s, 65°C for 30 s, and 72°C for 3.5 min, followed by 72°C for 7 min. The amplified products were separated by electrophoresis in a 1% (w/v) low melting point agarose Tris borate-EDTA gel and were observed by ethidium bromide staining. Three alternative splice products were consistently observed and each band was excised and sequenced. The tissue-specific distribution of the different splice products was assayed using RT-PCR on the same tissues used for RT-PCR analysis. The primers used for these reactions were *Mez2F10* (5'-CCATGTGAGAAGCAACCCTACAGC-3') and *Mez2R10* (5'-CCCAACCTGCAACATCAGATCTTCC-3'). The reaction conditions and cycling times were the same as those used for the RT-PCR reactions.

Received August 16, 2001; returned for revision November 28, 2001; accepted December 21, 2001.

LITERATURE CITED

- Aasland R, Stewart AF, Gibson T (1996) The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem Sci* 21: 87–88
- Akasaka T, Kanno M, Balling R, Mieza MA, Taniguchi M, Koseki H (1996) A role for *mel-18*, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* 122: 1513–1522
- Alkema MJ, van der Lugt NM, Bobeldijk RC, Berns A, van Lohuizen M (1995) Transformation of axial skeleton due to overexpression of *bmi-1* in transgenic mice. *Nature* 374: 724–727
- Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410: 120–124
- Brock HW, van Lohuizen M (2001) The Polycomb group: no longer an exclusive club? *Curr Opin Genet Dev* 11: 175–181
- Castle LA, Errampalli D, Atherton TL, Franzmann LH, Yoon ES, Meinke DW (1993) Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. *Mol Gen Genet* 241: 504–514
- Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ (1997) Fertilization-independent seed devel-

- opment in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 94: 4223–4228
- Cheung WL, Briggs SD, Allis CD (2000) Acetylation and chromosomal functions. Curr Opin Cell Biol 12: 326–333
- Core N, Bel S, Gaunt SJ, Aurrand-Lions M, Pearce J, Fisher A, Djabali M (1997) Altered cellular proliferation and mesoderm patterning in *Polycomb-M33*-deficient mice. Development 124: 721–729
- Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci USA 93: 8449–8454
- Franke A, DeCamillis M, Zink D, Cheng N, Brock HW, Paro R (1992) *Polycomb* and *polyhomeotic* are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. EMBO J 11: 2941–2950
- Galarneau L, Nourani A, Boudreau AA, Zhang Y, Heliot L, Allard S, Savari J, Lane WS, Stillman DJ, Cote J (2000) Multiple links between the NuA4 histone acetyltransferase complex and epigenetic control of transcription. Mol Cell 5: 927–937
- Gaut BS (2001) Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. Genome Res 11: 55–66
- Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94: 6809–6814
- Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G (1997) A *Polycomb*-group gene regulates homeotic gene expression in *Arabidopsis*. Nature 386: 44–51
- Gould A (1997) Functions of mammalian *Polycomb* group and trithorax group related genes. Curr Opin Gen Dev 7: 488–494
- Grossniklaus U, Vielle-Calzada JP, Hoepfner MA, Gagliano WB (1998) Maternal control of embryogenesis by *Medea*, a *Polycomb* group gene in *Arabidopsis*. Science 280: 446–450
- Helentjaris T (1995) Atlas of duplicated sequences. Maize Newslett 69: 67–81
- Holdeman R, Nehrt S, Strome S (1998) MES-2, a maternal protein essential for viability of the germline in *Caenorhabditis elegans*, is homologous to a *Drosophila* *Polycomb* group protein. Development 125: 2457–2467
- Jenuwein T (2001) Re-SET-ting heterochromatin by histone methyltransferases. Trends Cell Biol 11: 266–273
- Jones CA, Ng J, Peterson AJ, Morgan K, Simon J, Jones RS (1998) The *Drosophila* *esc* and *E(z)* proteins are direct partners in *Polycomb* group-mediated repression. Mol Cell Biol 18: 2825–2834
- Kelly WG, Fire A (1998) Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. Development 125: 2451–2456
- Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the MEDEA *Polycomb* gene in the *Arabidopsis* endosperm. Plant Cell 11: 1945–1952
- Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, Wells D, Katz A, Margossian L, Harada JJ, Goldberg RB et al. (1999) Control of fertilization-independent endosperm development by the *Medea* *Polycomb* gene in *Arabidopsis*. Proc Natl Acad Sci USA 96: 4186–4191
- Korf I, Fan Y, Strome S (1998) The *Polycomb* group in *Caenorhabditis elegans* and maternal control of germline development. Development 125: 2469–2478
- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410: 116–120
- Laible G, Wolf A, Dorn R, Reuter G, Nislow C, Leber-sorger A, Popkin D, Pillus L, Jenuwein T (1997) Mammalian homologues of the *Polycomb*-group gene *Enhancer of zeste* mediate gene silencing in *Drosophila* heterochromatin and at *S. cerevisiae* telomeres. EMBO J 16: 3219–3232
- Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A (2000) Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing *Arabidopsis* seeds. Proc Natl Acad Sci USA 97: 10637–10642
- Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM (1999) Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 96: 296–301
- Muller J, Gaunt S, Lawrence PA (1995) Function of the *Polycomb* protein is conserved in mice and flies. Development 121: 2847–2852
- Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292: 110–113
- Ng J, Hart CM, Morgan K, Simon JA (2000) A *Drosophila* ESC-E(Z) protein complex is distinct from other *Polycomb* group complexes and contains covalently modified ESC. Mol Cell Biol 20: 3069–3078
- Ng J, Li R, Morgan K, Simon J (1997) Evolutionary conservation and predicted structure of the *Drosophila* extra sex combs repressor protein. Mol Cell Biol 17: 6663–6672
- Ohad N, Margossian L, Hsu Y-C, Williams C, Repetti P, Fischer RL (1996) A mutation that allows endosperm development without fertilization. Proc Natl Acad Sci USA 93: 5319–5324
- Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL (1999) Mutations in FIE, a WD *Polycomb* group gene, allow endosperm development without fertilization. Plant Cell 11: 407–416
- Pal-Bhadra M, Bhadra U, Birchler JA (1997) Cosuppression in *Drosophila*: Gene silencing of *alcohol dehydrogenase* by *white-Adh* transgenes is *Polycomb* dependent. Cell 90: 479–490
- Pal-Bhadra M, Bhadra U, Birchler JA (1999) Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. Cell 99: 35–46
- Poux S, McCabe D, Pirrotta V (2001) Recruitment of components of *Polycomb* Group chromatin complexes in *Drosophila*. Development 128: 75–85
- Preuss D (1999) Chromatin silencing and *Arabidopsis* development: a role for *Polycomb* proteins. Plant Cell 11: 765–768
- Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis

- CD et al. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406: 593–599
- Sato T, Hayes PH, Denell RE (1984) Homeosis in *Drosophila*: maternal affect of the *Enhancer of Polycomb* locus and its interaction with *Polycomb* and related loci. *Dev Genet* 4: 185–198
- Sato T, Russell MA, Denell RE (1983) Homeosis in *Drosophila*: a new enhancer of *Polycomb* and related homeotic mutations. *Genetics* 105: 357–370
- Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* 28: 231–234
- Schumacher A, Magnuson T (1997) Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond. *Trends Genet* 13: 167–170
- Sewalt RG, van der Vlag J, Gunster MJ, Hamer KM, den Blaauwen JL, Satijn DP, Hendrix T, van Driel R, Otte AP (1998) Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. *Mol Cell Biol* 18: 3586–3595
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98: 37–46
- Shimono Y, Murakami H, Hasegawa Y, Takahashi M (2000) RET finger protein is a transcriptional repressor and interacts with *enhancer of polycomb* that has dual transcriptional functions. *J Biol Chem* 275: 39411–39419
- Simon J (1995) Locking in stable states of gene expression: transcriptional control during *Drosophila* development. *Curr Opin Cell Biol* 7: 376–385
- Spillane C, MacDougall C, Stock C, Kohler C, Vielle-Calzada JP, Nunes SM, Grossniklaus U, Goodrich J (2000) Interaction of the *Arabidopsis* polycomb group proteins FIE and MEA mediates their common phenotypes. *Curr Biol* 10: 1535–1538
- Stankunas K, Berger J, Ruse C, Sinclair DA, Randazzo F, Brock HW (1998) The *enhancer of polycomb* gene of *Drosophila* encodes a chromatin protein conserved in yeast and mammals. *Development* 125: 4055–4066
- Strutt H, Paro R (1997) The polycomb group protein complex of *Drosophila melanogaster* has different compositions at different target genes. *Mol Cell Biol* 17: 6773–6783
- Tachibana M, Sugimoto K, Fukushima T, Shinkai Y (2001) Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* 276: 25309–25317
- Tie F, Furuyama T, Prasad-Sinha J, Jane E, Harte PJ (2001) The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* 128: 275–286
- van der Lugt NM, Domen J, Linders K, van Roon M, Robanus-Maandag E, te Riele H, van der Valk M, Deschamps J, Sofroniew M, van Lohuizen M (1994) Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev* 8: 757–769
- van der Vlag J, Otte AP (1999) Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet* 23: 474–478
- van Lohuizen M, Tijms M, Voncken JW, Schumacher A, Magnuson T, Wientjens E (1998) Interaction of mouse polycomb-group (Pc-G) proteins Enx1 and Enx2 with Eed: indication for separate Pc-G complexes. *Mol Cell Biol* 18: 3572–3579
- Vielle-Calzada JP, Thomas J, Spillane C, Coluccio A, Hoepfner MA, Grossniklaus U (1999) Maintenance of genomic imprinting at the *Arabidopsis* *medea* locus requires zygotic DDM1 activity. *Genes Dev* 13: 2971–2982
- Xu L, Fong Y, Strome S (2001) The *Caenorhabditis elegans* maternal-effect sterile proteins, MES-2, MES-3, and MES-6, are associated in a complex in embryos. *Proc Natl Acad Sci USA* 98: 5061–5066
- Yadegari R, Kinoshita T, Lotan O, Cohen G, Katz A, Choi Y, Katz A, Nakashima K, Harada JJ, Goldberg RB et al. (2000) Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *Plant Cell* 12: 2367–2382